

0400
4/9/01

#4

OIPE

RAW SEQUENCE LISTING

DATE: 04/16/2001

PATENT APPLICATION: US/09/755,456

TIME: 11:59:14

Input Set : A:\156600.app

Output Set: N:\CRF3\04162001\I755456.raw

ENTERED

```

3 <110> APPLICANT: DELBAC, FREDERIC
4   DANCHIN, ANTOINE
5   VIVARES, CHRISTIAN
7 <120> TITLE OF INVENTION: MICROSPORIDIAN POLAR TUBE PROTEINS, NUCLEIC ACIDS
8   CODING FOR THESE PROTEINS AND THEIR APPLICATIONS
10 <130> FILE REFERENCE: 1566-00
12 <140> CURRENT APPLICATION NUMBER: 09/755,456
13 <141> CURRENT FILING DATE: 2001-01-05
15 <150> PRIOR APPLICATION NUMBER: PCT/FR99/01630
16 <151> PRIOR FILING DATE: 1999-07-06
18 <150> PRIOR APPLICATION NUMBER: FR 98/08692
19 <151> PRIOR FILING DATE: 1998-07-07
21 <160> NUMBER OF SEQ ID NOS: 10
23 <170> SOFTWARE: PatentIn Ver. 2.1
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 1830
27 <212> TYPE: DNA
28 <213> ORGANISM: Encephalitozoon cuniculi
30 <220> FEATURE:
31 <221> NAME/KEY: CDS
32 <222> LOCATION: (345)..(1529)
34 <400> SEQUENCE: 1
35 gaattcagat gcctcatacc ttgggattaa aaaattgatg ttcatttggtt atatatcctg 60
37 ggcggacagg ccggctcgta ttcttcagggt gtgtcgcccta cccagtgcac aggagggttcc 120
39 ggaggtgtct tggatggaaa gtaaggccat ttgtgggttc tcattccatgt catcgtcctt 180
41 ttcggctgtt tcaccaagat ccaattattc ctccaggact ttcaaccctc agaatggaaa 240
43 cagagatgaa actctctgtg caaatcgtag atatcgattg gagacattga aaccacggag 300
45 ttgaaataa aagtataaat acctccgaaa acgcagagtt taag atg aaa ggt att 356
46                                     Met Lys Gly Ile
47                                     1
49 tct aag atc ctc tct gcc tct att gcc ctg atg aag ttg gag aat gtc 404
50 Ser Lys Ile Leu Ser Ala Ser Ile Ala Leu Met Lys Leu Glu Asn Val
51   5          10          15          20
53 tat tca gca acc gca ctg tgc agc aat gca tat ggc cta act ccg gga 452
54 Tyr Ser Ala Thr Ala Leu Cys Ser Asn Ala Tyr Gly Leu Thr Pro Gly
55          25          30          35
57 caa cag ggt atg gct cag cag ccg tcg tat gtg ctg atc ccc agc acc 500
58 Gln Gln Gly Met Ala Gln Gln Pro Ser Tyr Val Leu Ile Pro Ser Thr
59          40          45          50
61 ccg gga acc ata gca aac tgt gca agc ggt tca cag gac aca tat tct 548
62 Pro Gly Thr Ile Ala Asn Cys Ala Ser Gly Ser Gln Asp Thr Tyr Ser
63          55          60          65
65 cct tct ccc gct gca ccc aca tct cca gtg act ccg ggg aaa act agc 596
66 Pro Ser Pro Ala Ala Pro Thr Ser Pro Val Thr Pro Gly Lys Thr Ser
67          70          75          80
69 gag aat gag aca tct cca tcg gct cct gca gaa gat gta gga aca tgc 644
70 Glu Asn Glu Thr Ser Pro Ser Ala Pro Ala Glu Asp Val Gly Thr Cys

```

RAW SEQUENCE LISTING

DATE: 04/16/2001

PATENT APPLICATION: US/09/755,456

TIME: 11:59:14

Input Set : A:\156600.app

Output Set: N:\CRF3\04162001\I755456.raw

```

71 85          90          95          100
73 aag att gcc gta ttg aag cac tgc gac gca cca gga aca aca tca ggg 692
74 Lys Ile Ala Val Leu Lys His Cys Asp Ala Pro Gly Thr Thr Ser Gly
75          105          110          115
77 acg aca cca ggg tca ggg cct tgt gaa acc cca gag cag caa cag cct 740
78 Thr Thr Pro Gly Ser Gly Pro Cys Glu Thr Pro Glu Gln Gln Gln Pro
79          120          125          130
81 ttg tca gtg atc tcc acc act cct gcc gta ccg gtg act gtg gag tct 788
82 Leu Ser Val Ile Ser Thr Thr Pro Ala Val Pro Val Thr Val Glu Ser
83          135          140          145
85 gca cag tct cca tct gtt gtg cca gtt gtt cct gtc gtt gct cac cac 836
86 Ala Gln Ser Pro Ser Val Val Pro Val Val Pro Val Val Ala His His
87          150          155          160
89 cag gca gtt cca ggc tac tac aac aat gga aca tcc ggt att cct gga 884
90 Gln Ala Val Pro Gly Tyr Tyr Asn Asn Gly Thr Ser Gly Ile Pro Gly
91 165          170          175          180
93 cag caa cag atc ctt tct ggc act ctt ccc cca gga gcc act ttg tgt 932
94 Gln Gln Gln Ile Leu Ser Gly Thr Leu Pro Pro Gly Ala Thr Leu Cys
95          185          190          195
97 cag gga cag gcc atg cct agc act cct gga cag caa cag atc ctt tct 980
98 Gln Gly Gln Ala Met Pro Ser Thr Pro Gly Gln Gln Gln Ile Leu Ser
99          200          205          210
101 ggc act ctt ccc cca ggg gtc act ttg tgt cag gga cag gcc acg cct 1028
102 Gly Thr Leu Pro Pro Gly Val Thr Leu Cys Gln Gly Gln Ala Thr Pro
103          215          220          225
105 agc act cct ggg cag caa cag gtc ctt tct ggc act ctt ccc cca gga 1076
106 Ser Thr Pro Gly Gln Gln Gln Val Leu Ser Gly Thr Leu Pro Pro Gly
107          230          235          240
109 gtc act ttg tgt cag gga cag gcc acg cct agc act cct ggg cag caa 1124
110 Val Thr Leu Cys Gln Gly Gln Ala Thr Pro Ser Thr Pro Gly Gln Gln
111 245          250          255          260
113 cag gtc ctt tct ggc acc ctt ctc cca gga gcc act ttg tgt cag gat 1172
114 Gln Val Leu Ser Gly Thr Leu Leu Pro Gly Ala Thr Leu Cys Gln Asp
115          265          270          275
117 caa ggt atg cct gga aca tcc gga gtt cct gga cag cag gga cag tct 1220
118 Gln Gly Met Pro Gly Thr Ser Gly Val Pro Gly Gln Gln Gly Gln Ser
119          280          285          290
121 agt gga cag tgt tgt gcc cct cag att cca aac cct gtc atg ccg cca 1268
122 Ser Gly Gln Cys Cys Ala Pro Gln Ile Pro Asn Pro Val Met Pro Pro
123          295          300          305
125 tcc atg aac att agt gga aat ggg tat cct tct tct acc gca tac agc 1316
126 Ser Met Asn Ile Ser Gly Asn Gly Tyr Pro Ser Ser Thr Ala Tyr Ser
127          310          315          320
129 cca aac ctc gga tca ctg gga tcc tgt gtt gac ata cag aag acg ggg 1364
130 Pro Asn Leu Gly Ser Leu Gly Ser Cys Val Asp Ile Gln Lys Thr Gly
131 325          330          335          340
133 ggg aca tcc tgc gag caa aaa ccc gag aag tcc gcc acg cag tat gcc 1412
134 Gly Thr Ser Cys Glu Gln Lys Pro Glu Lys Ser Ala Thr Gln Tyr Ala
135          345          350          355

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/755,456

DATE: 04/16/2001

TIME: 11:59:14

Input Set : A:\156600.app

Output Set: N:\CRF3\04162001\I755456.raw

```

137 atg gag gcc tgt gca aca cca aca cca acg gtt att ata ggc aac agc 1460
138 Met Glu Ala Cys Ala Thr Pro Thr Pro Thr Val Ile Ile Gly Asn Ser
139          360          365          370
141 gag tat ctt gtt gga cca gga atg tac aat gca att aac tct cca tgc 1508
142 Glu Tyr Leu Val Gly Pro Gly Met Tyr Asn Ala Ile Asn Ser Pro Cys
143          375          380          385
145 aac act gct gtc caa tgc tgc taggctaaaa taaaacgagt ttaatcttct 1559
146 Asn Thr Ala Val Gln Cys Cys
147          390          395
149 ttttcttcgg tcttttggaa cgttggatgg ggatggagga gtctatgggc tgaagtgaag 1619
151 tgccaacact tcttctgccc aagaacacat tcggatgttc ttctgtggc caggagtttg 1679
153 gtaacaggat tcccggagga tttagcagcc ttggagtacc atgattgaat cagtattaaa 1739
155 cttctcaaat tattttattc tttctgtttt atatcccgag ccaatctgag aagaatgcct 1799
157 cgaattcaag ctcccttaga agtgtgggat c 1830
160 <210> SEQ ID NO: 2
161 <211> LENGTH: 1740
162 <212> TYPE: DNA
163 <213> ORGANISM: Encephalitozoon cuniculi
165 <220> FEATURE:
166 <221> NAME/KEY: CDS
167 <222> LOCATION: (458)..(1288)
169 <400> SEQUENCE: 2
170 aagcttctga acaagcgcta accctcttctc agaatatata aagcaatcca tacaacttct 60
172 ccatccatcc cgggtgctgtt tctttggagg caaaacagag gaggtggcga tatcgatggt 120
174 gcatccataa tatatacaag acactccagg ctgcaactga atcaacacac tccatcccct 180
176 caggaagtcg gtaaaacttc cttgaaaata gccaatggat gtctccaggc ttataccat 240
178 gcacagctat atcttggcct gaagtgcact ttcaggtggg gctttgttac attgcggtgt 300
180 tttggattac ctgatataat ttgttaccga ctgagtcaag tcgaaaccag tagtccgcag 360
182 atttctaaca gagaggaaag actggaggta atttgtggct ttgaaacat gcacagcaaa 420
184 ataaaaatata aaagaagcct tttgcacact accaaag atg ttg tta ctt ctc gcc 475
185                                     Met Leu Leu Leu Leu Ala
186                                     1           5
188 ata act gct gtt gtt agc gcc acg atg gtc cat cct tca gct gtt gtt 523
189 Ile Thr Ala Val Val Ser Ala Thr Met Val His Pro Ser Ala Val Val
190          10          15          20
192 cca cag ccc gca gca cct ctc cat gtc gtt ccc cca cag cag caa atg 571
193 Pro Gln Pro Ala Ala Pro Leu His Val Val Pro Pro Gln Gln Gln Met
194          25          30          35
196 ggc atg gtt aac gga tgc acc agc aag aaa cta gag ggt gca gaa ata 619
197 Gly Met Val Asn Gly Cys Thr Ser Lys Lys Leu Glu Gly Ala Glu Ile
198          40          45          50
200 atg aga agg aac atg att gag tgc cag aaa aga agc tcg gag gca aca 667
201 Met Arg Arg Asn Met Ile Glu Cys Gln Lys Arg Ser Ser Glu Ala Thr
202          55          60          65          70
204 aag gcg atg att gaa agg gca aat gaa aag gct gta gaa tca ttc aac 715
205 Lys Ala Met Ile Glu Arg Ala Asn Glu Lys Ala Val Glu Ser Phe Asn
206          75          80          85
208 aag gaa gtt agc aaa gga cct agc caa aag gat gga ggc cag tgc ata 763
209 Lys Glu Val Ser Lys Gly Pro Ser Gln Lys Asp Gly Gly Gln Cys Ile

```

RAW SEQUENCE LISTING

DATE: 04/16/2001

PATENT APPLICATION: US/09/755,456

TIME: 11:59:14

Input Set : A:\156600.app

Output Set: N:\CRF3\04162001\I755456.raw

```

210          90          95          100
212 gaa aaa gct gta caa ggt acc gat agg tgt att ctc gct gga ata atc 811
213 Glu Lys Ala Val Gln Gly Thr Asp Arg Cys Ile Leu Ala Gly Ile Ile
214          105          110          115
216 gat aag gcg gtg aac aag cgc aag tac aga atc tca gat gtg gag aac 859
217 Asp Lys Ala Val Asn Lys Arg Lys Tyr Arg Ile Ser Asp Val Glu Asn
218          120          125          130
220 agc acc tcg ctc tac aga gga gac aag cta att gcc cta att gtc aat 907
221 Ser Thr Ser Leu Tyr Arg Gly Asp Lys Leu Ile Ala Leu Ile Val Asn
222 135          140          145          150
224 gtc gac tat ggg ctg cag ccg atc act aag cca aag aag aag aag tcc 955
225 Val Asp Tyr Gly Leu Gln Pro Ile Thr Lys Pro Lys Lys Lys Ser
226          155          160          165
228 aag ata atg gcg aat ctc cct cag ccg aag aga gag atg tat ttc aac 1003
229 Lys Ile Met Ala Asn Leu Pro Gln Pro Lys Arg Glu Met Tyr Phe Asn
230          170          175          180
232 caa atc ggt cag ctt gtt gga gca aga gga acg ttc ccc cag gaa aac 1051
233 Gln Ile Gly Gln Leu Val Gly Ala Arg Gly Thr Phe Pro Gln Glu Asn
234          185          190          195
236 aag gag gac tgc aag cct tgt gag ggt ccc aag aag act gtt gaa act 1099
237 Lys Glu Asp Cys Lys Pro Cys Glu Gly Pro Lys Lys Thr Val Glu Thr
238          200          205          210
240 act tct gag aaa tgt aat ctt ggg tgc gag ctt aaa gga aca tct gct 1147
241 Thr Ser Glu Lys Cys Asn Leu Gly Cys Glu Leu Lys Gly Thr Ser Ala
242 215          220          225          230
244 ctg ata agc aag gcc ata cag aag aag gaa gtc aag gac acg aag gaa 1195
245 Leu Ile Ser Lys Ala Ile Gln Lys Lys Glu Val Lys Asp Thr Lys Glu
246          235          240          245
248 ggg gag aaa agt gca agc cag gac tct gat ggc gag ggc act gct gag 1243
249 Gly Glu Lys Ser Ala Ser Gln Asp Ser Asp Gly Glu Gly Thr Ala Glu
250          250          255          260
252 gat gcg gaa gta cag caa cct tct gcg gac ggc gag ggt cta gag 1288
253 Asp Ala Glu Val Gln Gln Pro Ser Ala Asp Gly Glu Gly Leu Glu
254          265          270          275
256 taatttttaa attaaaatct ccctggattg aatcttcaag tgcttttgtg aaagactttg 1348
258 ggaacatttc gtgaaggcta acataaattg ttaatctcag gtcactcgat ggaatagtca 1408
260 attcgatattt cctttccttg gatggtctgc cccaccagcc tgttcctggc agttatcgca 1468
262 tcgtcgacag agtcaaaactg aacgaatcca tatcctttgg acatcttctt gtattggtcg 1528
264 tagactatta ctaccgata gttcagtatc tcactgatcc tctccttgag aaggtctcta 1588
266 acgtcgtctt cggttatgtg tgctccagc ccaaatatcc ctatcgccct ggaggagac 1648
268 ccgtttctct ttgctttaag tgcatatctt tcgtttttat aggagcttg atctgttct 1708
270 tcgtatcccc ttgtcgggcg ctccacctcg ag 1740
273 <210> SEQ ID NO: 3
274 <211> LENGTH: 1116
275 <212> TYPE: DNA
276 <213> ORGANISM: Encephalitozoon intestinalis
278 <220> FEATURE:
279 <221> NAME/KEY: CDS
280 <222> LOCATION: (1)..(1113)

```

RAW SEQUENCE LISTING

DATE: 04/16/2001

PATENT APPLICATION: US/09/755,456

TIME: 11:59:14

Input Set : A:\156600.app

Output Set: N:\CRF3\04162001\I755456.raw

```

282 <400> SEQUENCE: 3
283 atg aaa ggt att tct aag gtt ctc tca gcc tct att gtc cta atg aag 48
284 Met Lys Gly Ile Ser Lys Val Leu Ser Ala Ser Ile Val Leu Met Lys
285 1 5 10 15
287 ttg aag ggt gtc tat tct aca act gtg ctg tgt gga gat tca aca caa 96
288 Leu Lys Gly Val Tyr Ser Thr Thr Val Leu Cys Gly Asp Ser Thr Gln
289 20 25 30
291 gga ctg cag ggc aca acc caa ccg tca tat gtg ctg gtt cct agt gca 144
292 Gly Leu Gln Gly Thr Thr Gln Pro Ser Tyr Val Leu Val Pro Ser Ala
293 35 40 45
295 cca gag aca ata gcc aac tgt gga tac agt cca cag aac atg tat gtc 192
296 Pro Glu Thr Ile Ala Asn Cys Gly Tyr Ser Pro Gln Asn Met Tyr Val
297 50 55 60
299 cct tct act cct act acc atg cct tcc aca gtg cca ggc aca act ggt 240
300 Pro Ser Thr Pro Thr Thr Met Pro Ser Thr Val Pro Gly Thr Thr Gly
301 65 70 75 80
303 gag agc gag aca cct act tct cca aca tca tct cct aca gag gat gtg 288
304 Glu Ser Glu Thr Pro Thr Ser Pro Thr Ser Ser Pro Thr Glu Asp Val
305 85 90 95
307 gga aca tgc aag att gct gtt gta aag cat tgt gat gca cca gga aca 336
308 Gly Thr Cys Lys Ile Ala Val Val Lys His Cys Asp Ala Pro Gly Thr
309 100 105 110
311 tca tca aca cct tgc gaa ccg gaa cag act ttg gcc ccc tct cag cca 384
312 Ser Ser Thr Pro Cys Glu Pro Glu Gln Thr Leu Ala Pro Ser Gln Pro
313 115 120 125
315 gta gca gct aca att gcc aca cca ctg gtt gtt gct tct gtg cag acg 432
316 Val Ala Ala Thr Ile Ala Thr Pro Leu Val Val Ala Ser Val Gln Thr
317 130 135 140
319 ccg caa gca gct gtt acc atc ctt act cca aag gcc gtc tct gcc cag 480
320 Pro Gln Ala Ala Val Thr Ile Leu Thr Pro Lys Ala Val Ser Ala Gln
321 145 150 155 160
323 ccg gca acc atc att tct cca ttc aac cag gca cca ggc tac tac aat 528
324 Pro Ala Thr Ile Ile Ser Pro Phe Asn Gln Ala Pro Gly Tyr Tyr Asn
325 165 170 175
327 agt gca att ccc ggg caa ata ctt aca ggt aat gtt ctc tct cca agt 576
328 Ser Ala Ile Pro Gly Gln Ile Leu Thr Gly Asn Val Leu Ser Pro Ser
329 180 185 190
331 gcc tct tct tgc caa gtg gtg ccc gga aca aca gga agc tcc acc ccc 624
332 Ala Ser Ser Cys Gln Val Val Pro Gly Thr Thr Gly Ser Ser Thr Pro
333 195 200 205
335 cag cag cta cca ggc gct gtt tca tct gga acc att cct tgc caa ata 672
336 Gln Gln Leu Pro Gly Ala Val Ser Ser Gly Thr Ile Pro Cys Gln Ile
337 210 215 220
339 gta cag gga act caa agt agc gga aac acc cct gga cag caa ttc ttg 720
340 Val Gln Gly Thr Gln Ser Ser Gly Asn Thr Pro Gly Gln Gln Phe Leu
341 225 230 235 240
343 ccg gga atc gtt cct gtt gga agc ctc cag ccg gat caa gct act tct 768
344 Pro Gly Ile Val Pro Val Gly Ser Leu Gln Pro Asp Gln Ala Thr Ser
345 245 250 255

```

VERIFICATION SUMMARY

DATE: 04/16/2001

PATENT APPLICATION: US/09/755,456

TIME: 11:59:15

Input Set : A:\156600.app

Output Set: N:\CRF3\04162001\I755456.raw